# A Survey of Multimodel Large Language Models

## Abstract

With the widespread application of the Transformer architecture in various modalities, including vision, the technology of large language models is evolving from a single modality to a multi-modal approach. It is foreseeable that multi-modal large language models will become one of the most powerful tools for humanity in the future. To proactively adapt to technological changes and grasp the development trends of multi-modal large language models, this article provides a comprehensive review of such models. The review is concise yet thorough, covering aspects such as the background, significance, definition, characteristics, and application scenarios of multi-modal large language models. Emphasis is placed on elucidating the developmental trajectory of these models, introducing seven different model architectures within the multi-modal context. The aim is to further advance the field of multi-modal large language models.